Clustering micropollutants based on initial biotransformations for improved prediction of micropollutant removal during conventional activated sludge treatment

Wang, Yuxin ; Fenner, Kathrin ; Helbling, Damian E

Abstract: The lack of fundamental insights on the fate of micropollutants during conventional activated sludge treatment in wastewater treatment plants (WWTPs) presents one of the biggest challenges in optimizing their removal. To address this challenge, we designed a study to identify the drivers of micropollutant removal in WWTPs. We calculated the removal efficiency of micropollutants across the activated sludge process of full-scale wastewater treatment plants from literature-reported data. Our final dataset consisted of 529 independent observations for 84 micropollutants along with a set of associated activated sludge process parameters. We used the Eawag pathway prediction system to predict initial biotransformations for each of the 84 micropollutants and hierarchical clustering to group chemicals based on similarities in their predicted initial sets of biotransformations. We then applied stability selection to generate well-performing models that can be interpreted to uncover the key factors contributing to the range of removal efficiencies calculated for each cluster of micropollutants. The key factors considered in stability selection included six physicochemical properties of the micropollutants and six activated sludge process parameters. The sludge-water partitioning coefficient, molecular weight, solids retention time, influent micropollutant concentration, and dissolved oxygen levels were consistently identified as key factors of micropollutant removal for clusters of micropollutants that undergo certain types of initial biotransformations. Our findings highlight the importance of considering initial biotransformations when evaluating micropollutant removal and identify important process parameters that determine the fate of micropollutants during activated sludge treatment.

DOI: https://doi.org/10.1039/C9EW00838A

TITLE: Clustering micropollutants based on initial biotransformations for improved prediction of micropollutant removal during conventional activated sludge treatment

AUTHORS: Yuxin Wang [a], Kathrin Fenner [b,c,d], Damian E. Helbling [a,*]

AFFILIATIONS:

[a] School of Civil and Environmental Engineering, Cornell University, Ithaca, NY, USA

[b] Eawag, Überlandstrasse 133, 8600, Dübendorf, Swizerland

[c] Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, 8092, Zürich, Switzerland

[d] Department of Chemistry, University of Zürich, 8057, Zürich, Switzerland

***,** Corresponding author**:** Damian E. Helbling, School of Civil and Environmental Engineering, Cornell University, 220 Hollister Hall, Ithaca, NY, 14853, USA. Email: damian.helbling@cornell.edu. Tel: +1 607 255 5146. Fax: +1 607 255 9004.

WATER IMPACT STATEMENT:

Despite decades of research, the main drivers of micropollutant removal during wastewater treatment remain unknown. We developed multivariable models to explain variability in removal efficiencies among micropollutants that undergo the same types of biotransformations. Our models predict removal efficiencies for micropollutants that undergo the same types of biotransformations and can be interpreted to identify the main drivers of micropollutant removal.

This document is the accepted manuscript version of the following article:
Wang, Y., Fenner, K., & Helbling, D. E. (2020). Clustering micropollutants based on initial biotransformations for improved prediction of micropollutant removal during conventional activated sludge treatment. Environmental Science: Water Research and Technology. https://doi.org/10.1039/C9EW00838A

Abstract

The lack of fundamental insights on the fate of micropollutants during conventional activated sludge treatment in wastewater treatment plants (WWTPs) presents one of the biggest challenges in optimizing their removal. To address this challenge, we designed a study to identify the drivers of micropollutant removal in WWTPs. We calculated the removal efficiency of micropollutants across the activated sludge process of full-scale wastewater treatment plants from literature-reported data. Our final dataset consisted of 529 independent observations for 84 micropollutants along with a set of associated activated sludge process parameters. We used the Eawag Pathway Prediction System to predict initial biotransformations for each of the 84 micropollutants and hierarchical clustering to group chemicals based on similarities in their predicted initial sets of biotransformations. We then applied stability selection to generate well-performing models that can be interpreted to uncover the key factors contributing to the range of removal efficiencies calculated for each cluster of micropollutants. The key factors considered in stability selection included six physicochemical properties and six activated sludge process parameters. The sludge-water partitioning coefficient, molecular weight, solids retention time, influent micropollutant concentration, and dissolved oxygen levels were consistently identified as key factors of micropollutant removal for clusters of micropollutants that undergo certain types of initial biotransformations. Our findings highlight the importance of considering initial biotransformations when evaluating micropollutant removal and identify important process parameters that determine the fate of micropollutants during activated sludge treatment.

1. Introduction

Effluent from conventional wastewater treatment plants (WWTPs) is a major source of micropollutants in the aquatic environment.[1–3] As WWTPs are final treatment steps prior to the release of micropollutants into the environment,[4] a better understanding of how process parameters impact the fate of micropollutants during wastewater treatment is critical. Individual WWTPs remove diverse types of micropollutants to varying extents, and individual micropollutants are removed to varying extents among different WWTPs.[5] This variability is often attributed to the physicochemical properties of the micropollutants and variability in process parameters.[6,7] However, a general consensus on the main drivers of micropollutant removal during wastewater treatment is lacking,[8,9] which could be because the critical parameter is unknown or because a single parameter is unable to explain the removal of structurally different micropollutants.[10] Some suggest that accurate prediction of micropollutant removal in WWTPs may not be possible.[11,12]

We suspect that there are two main reasons that no general consensus has emerged on the main drivers of micropollutant removal during wastewater treatment. First, much of the previous literature exploring the parameters that influence micropollutant removal in WWTPs is based on data that aggregates measurements collected from different processes.[1,3,11,13–16] However, micropollutant removal mechanisms vary significantly among treatment processes within a WWTP. For example, adsorption to solids or volatilization are expected to be the dominant removal mechanism during primary treatment, whereas a combination of biotransformation and adsorption to activated sludge flocs are the dominant removal mechanisms during secondary treatment.[17] We hypothesize that accurate and robust predictions of micropollutant removal during wastewater treatment can be achieved only by isolating specific removal mechanisms and

61 exploring them independently. Second, previous studies that explore the parameters that
62 influence micropollutant biotransformation in WWTPs[9,18] have focused on micropollutants that
63 likely undergo very different enzyme-catalyzed biotransformations during wastewater
64 treatment.[19,20] The performance of predictive models developed in these studies has been
65 relatively poor, with one recently reported model accounting for only 17% of the observed
66 variability in the micropollutant removal data.[9] We surmise that the poor performance of
67 previously reported models is the result of aggregating micropollutants with disparate chemical
68 structures that most likely undergo very different biotransformations during wastewater
69 treatment. We therefore hypothesize that accurate and robust predictions of micropollutant
70 removal during wastewater treatment can be achieved only for groups of micropollutants that
71 undergo the same types of enzyme-catalyzed biotransformations, as has also been suggested by
72 some recent experimental work.[21,22]

73 The lack of fundamental insights on the fate of micropollutants in WWTPs presents one
74 of the biggest challenges in optimizing their removal during wastewater treatment. To address
75 this challenge and to test our hypotheses, we designed a study that aimed to identify the drivers
76 of micropollutant removal in WWTPs. We focused our study on biotransformations during
77 activated sludge treatment as a means to isolate a single, but major, micropollutant removal
78 mechanism (*i.e.*, biotransformations). We then developed a novel approach to identify the drivers
79 of micropollutant removal within groups of micropollutants that undergo the same types of
80 enzyme-catalyzed biotransformations. We first collected a comprehensive set of literature-
81 reported data on micropollutant removal in conventional activated sludge systems. We then used
82 the Eawag Pathway Prediction System (Eawag-PPS) to predict initial biotransformations for
83 each of the micropollutants in the dataset. We used hierarchical clustering to aggregate

84 micropollutants based on similarities in their putative initial biotransformation profiles. We
85 finally performed a multivariable analysis to identify the physicochemical properties or process
86 parameters that determine the variable extents of removal among the micropollutants contained
87 within the resulting micropollutant clusters. Our findings validate our hypotheses and allow us to
88 identify a set of physicochemical properties and process parameters that accurately and robustly
89 predict the removal of certain groups of micropollutants during wastewater treatment.

90 **2. METHODS**

91 *2.1 Literature review and data collection*

92 We performed a literature review to identify studies that report measured concentrations of
93 micropollutants in samples collected from the influent and effluent of conventional activated
94 sludge systems. To be as comprehensive as possible, we first identified all review papers that
95 were returned on a Web of Science search using the search terms: "micropollutant*" OR
96 "emerging contaminant*" OR "pharmaceuticals" OR "pharmaceuticals and personal care
97 products") AND ("removal" OR "elimination") AND (“activated sludge” OR “conventional
98 activated sludge*”) AND ("wastewater treatment plant*" OR “sewage treatment plant*”). We
99 reviewed approximately 400 manuscripts that were identified in this search and collected data
100 from the 37 manuscripts that reported dissolved micropollutant concentrations, removal
101 efficiencies, or biodegradation rate constants from data collected exclusively at the influent to the
102 activated sludge process (*i.e.*, the effluent of the primary clarifier) and the effluent of the
103 secondary clarifier along with associated process parameters. We limited the data collection to
104 these 37 studies as a means to isolate the removal processes occurring during activated sludge
105 treatment and to eliminate, to the extent possible, variability in calculated removal efficiencies
106 resulting from variable concentrations, variable flow, and complicated mixing patterns[23,24] that

can be equalized during primary clarification.[25] The sampling schemes implemented in the 37 studies included a mixture of 24-hour composite sampling and grab sampling, though several studies lacked sufficient detail to fully characterize the sampling scheme. To generate a self-consistent dataset, we converted reported micropollutant concentrations to removal efficiency ($R\%$) using Equation 1.

$$R\% = \left[1 - \frac{C_{eff}}{C_{inf}}\right] \times 100 \qquad \text{Equation 1}$$

where $R\%$ is the micropollutant removal efficiency and C_{inf} and C_{eff} are the influent and effluent concentration [$\mu g \cdot L^{-1}$]. Micropollutants with reported influent concentrations that were below the limit of quantification (LOQ) were discarded. Micropollutants with effluent concentrations that were below the LOQ were assigned values equal to half the reported LOQ to be consistent with how $R\%$ was calculated in the majority of the studies that reported $R\%$.[26] Reported biodegradation rate constants were converted to $R\%$ by using Equation 2 as previously described.[4,7]

$$R\% = [1 - e^{-k_{bio} \times MLSS \times HRT}] \times 100 \qquad \text{Equation 2}$$

where k_{bio} is the reported second-order rate constant [$L \cdot g^{-1} \cdot hr^{-1}$], $MLSS$ is the reported mixed liquor suspended solids concentration [$g \cdot L^{-1}$], and HRT is the reported hydraulic retention time [hr]. Data from studies that reported k_{bio} but did not report $MLSS$ or HRT were discarded. Citations to the 37 manuscripts are provided in **Table S1** of the **Electronic Supplementary Information (ESI)**. The parameters collected as associated process metadata are summarized in **Table S2** of the **ESI**.

Six physicochemical properties for each micropollutant that may be important for influencing biotransformation were also recorded. We used EPISUITE and the Open Chemistry

Database from the National Institutes of Health to obtain experimental (preferred) or estimated values for the octanol-water partitioning coefficient (K_{ow}), water solubility (S), and Henry's Law constant (HLC). The Marvin software was used to record the overall charge of each micropollutant at pH 7 (*charge*); neutral micropollutants were assigned an overall charge of zero.[27] We collected activated sludge-water distribution coefficients (K_d) for each micropollutant from the literature. If K_d values were not identified in the literature, we estimated K_d values from the organic carbon-water partition coefficient (K_{oc}) using a constant value of 0.37 for the organic carbon fraction of activated sludge (f_{oc}).[28,29] Experimental (preferred) or estimated values for K_{oc} were obtained from EPISUITE. Finally, molecular weight (MW) was included as a metric of molecule size.

2.2 Data Filtering to Generate Final Dataset for Analysis

Our complete dataset contains 1,076 entries related to 169 micropollutants. It is not uncommon to encounter negative values for *R%* in the literature,[7,30,31] a phenomenon often attributed to poorly designed sampling strategies,[32] desorption of micropollutants from fecal particles or return activated sludge,[33] or deconjugation of metabolites during activated sludge treatment.[33,34] Because we could not disentangle these processes, we excluded all entries with negative *R%* values, which reduced our dataset to 781 entries related to 155 micropollutants. Also, because we are primarily interested in identifying the physicochemical properties or process parameters that influence *biotransformation* of micropollutants in activated sludge processes, we restricted our analysis to only those micropollutants with a reported or estimated K_d value of less than 500 $L \cdot kg^{-1}$. Micropollutants with K_d values greater than 500 $L \cdot kg^{-1}$ would be removed by a combination of biotransformation and adsorption to sludge, making it difficult to isolate the

149 factors influencing biotransformation.[35–37] The final database contains 529 entries related to 84
150 micropollutants.
151 *2.3 Data processing – multiple imputation*
152 Our dataset contains a significant amount of missing values because each study reported a
153 varying number of process parameters. To address this issue, multiple imputation was employed
154 with the *MICE* package in R using the predicted mean matching imputation method to fill in
155 missing values as previously described.[38,39] Values were imputed using a regression between the
156 process parameters and *R%* plus a random term drawn from the regression residuals,[40,41] an
157 approach that accounts for uncertainty in the imputation and provides valid variance estimates.[42–
158 [44] We conservatively applied multiple imputation to generate ten complete datasets that contain
159 plausible values for each of the missing values for process parameters that contained less than 60%
160 missing values, a threshold that was established using previously described criteria.[38]
161 *2.4 Data processing – hierarchical clustering*
162 One hypothesis of this study is that the *R%* of micropollutants that undergo the same initial
163 biotransformations can be predicted by a single set of physicochemical and process parameters.
164 To test this hypothesis, we first predicted the initial biotransformations for each of the 84
165 micropollutants using the Eawag pathway prediction system (Eawag-PPS). The Eawag-PPS uses
166 a set of biotransformation rules (btrules) based on fragment recognition to predict plausible
167 biotransformations for organic chemicals.[45] We predicted initial biotransformations in four ways:
168 (1) relative reasoning, neutral likelihood cutoff; (2) no relative reasoning, neutral likelihood
169 cutoff; (3) relative reasoning, likely likelihood cutoff; and (4) no relative reasoning, likely
170 likelihood cutoff. Relative reasoning allows certain biotransformations to have priority over
171 others based on experimental observations; likelihood cutoffs consider the likelihood of a

particular biotransformation based on expert knowledge.[46] We built matrices containing the number of times each micropollutant triggered a given btrule as an initial biotransformation step. We then used hierarchical clustering to identify groups of micropollutants that undergo similar sets of predicted initial biotransformations. Micropollutant clusters were determined by hierarchical clustering with the *hclust* function in R, using Ward's agglomeration method and Euclidean distance matrices.[47]

2.5 Generalized additive modelling and stepwise regression

The final dataset for multivariable modelling contained *R%* as a response variable along with six physicochemical properties (molecular weight, *MW*, solubility, *S*, octanol-water partitioning coefficient, K_{ow}, K_d, *charge*, and *HLC*) and six process parameters (population served, *P*; flow rate, *Q*; hydraulic retention time, *HRT*; solids retention time, *SRT*; micropollutant concentration in the influent, C_{inf}; and nitrification/denitrification, *NDN*; see **Table S2** for definitions and units) for which there was sufficient data for multiple imputation as predictor variables. Prior to analysis, all of the predictor variables (except *charge*) were log transformed to increase the linearity of the response and to transform the data to an approximately normal distribution.[9] A generalized additive model (GAM) was used to develop shape functions for each predictor variable (as derived from the quantitative relationships) as shown in **Figure S1**, which generated a multivariable model provided as Equation 3; predictor variables that exhibit monotonic shape functions are included in Equation 3 as linear additive terms whereas predictor variables that exhibit non-monotonic shape functions are included in Equation 3 as non-linear additive terms. Equation 3 was then used with stepwise regression[38] to evaluate the prediction of *R%* on: (1) the complete dataset; (2) a set of individual micropollutants defined as "seed" micropollutants; and (3) on micropollutant clusters identified by means of hierarchical clustering.

$$LogR\% = LogK_d + LogMW + LogK_{ow} + (LogK_{ow})^2 + LogS + (LogS)^2 + Charge + Charge^2 + LogHLC + LogP + LogQ + LogHRT + (LogHRT)^2 + LogSRT + LogC_{inf} + (LogC_{inf})^2 + LogNDN$$

Equation 3

2.6 Stability selection to identify important predictors

We next aimed to generate well-performing models to describe the variability in the response variables contained within each micropollutant cluster as a means to identify the key physicochemical or process parameters that determine their removal during activated sludge treatment. We implemented a stability selection approach that combines stepwise regression with bootstrap subsampling to generate robust model scenarios as has been previously described.[38,48] Briefly, we used bootstrap resampling with replacement to generate 1,000 datasets for each of the ten imputed datasets for each micropollutant cluster. The sample size for bootstrap resampling was selected to be equal to the original population size of the data for each micropollutant cluster.[49–51] We then implemented stepwise regression to identify models that generate the highest accuracy while incorporating the fewest number of predictor variables. We applied stepwise regression to each of the bootstrapped datasets to evaluate how frequently each of the predictor variables was selected into well-performing models. We defined well-performing stepwise regression models as those that performed as well as or better than the GAM model as measured by adjusted R^2 values. We used the frequency that each predictor variable was selected into well-performing models as a metric of how generally important that predictor variable is in determining *R%*.[48] A variance inflation factor (VIF) was applied to evaluate predictor collinearity, and predictors with a VIF value larger than 10 were excluded from the analysis.[52] We constructed multilinear models using the final set of predictors for each cluster for each of

214 the 10,000 bootstrapped subsamples and recorded the averaged adjusted R^2 as a metric of model
215 performance. We constructed GAM shape functions using the final set of predictors for each
216 cluster for all of the imputed datasets to evaluate the contribution of each predictor to *R%*.

217 **3. RESULTS AND DISCUSSION**

218 *3.1 Analysis of complete dataset*

219 The final dataset contains 529 independent values of *R%* for 84 micropollutants whose
220 removal in activated sludge processes is expected to be dominated by biotransformation. To the
221 best of our knowledge, this is the only dataset that has been compiled with a focus on
222 micropollutant biotransformations occurring exclusively in conventional activated sludge
223 processes. Boxplots summarizing *R%* for each of the micropollutants with more than four
224 independent observations are provided in **Figure 1**. In general, the data in **Figure 1** confirms our
225 expectation that individual micropollutants are removed to varying extents among different
226 WWTPs. We note that the *R%* values for some of the poorly removed micropollutants (*e.g.*,
227 carbamazepine) may be skewed due to the exclusion of all data entries with negative *R%* values.
228 We also performed a GAM and stepwise regression on *R%* as a response variable with six
229 physicochemical properties and six process parameters as predictor variables for each of the ten
230 imputed datasets. The analysis resulted in adjusted R^2 values between 0.15 and 0.19 among the
231 ten imputed datasets, and the results of this analysis for the imputed dataset that yielded the
232 highest adjusted R^2 value are presented in **Figure 2**.

233 The data presented in **Figure 1** and the results presented in **Figure 2** lead to two
234 important insights. First, the reported *R%* values of the 84 micropollutants are highly variable;
235 this variability reflects differences among the micropollutants and differences for individual
236 micropollutants that we expect to be explained by a common set of predictors. Second, similar to

237 previous studies,[9,18] our GAM and stepwise regression perform poorly with respect to explaining
238 the observed variability in *R%* values among all micropollutants. This is not an unexpected result;
239 because micropollutants are biotransformed in activated sludge processes in different ways, we
240 do not expect that a single set of physicochemical properties and process parameters will be
241 sufficient to predict *R%* for all micropollutants. Rather, we expect that only micropollutants that
242 undergo the same types of enzyme-catalyzed biotransformations will have *R%* values that can be
243 explained by a common set of predictors.

244 *3.2 Clustering micropollutants based on predicted biotransformations*

245 Because our database has isolated a set of micropollutants whose removal in activated
246 sludge processes is expected to be dominated by biotransformation, we clustered micropollutants
247 together based on their potential to undergo similar sets of initial biotransformations. We used
248 four prediction scenarios in the Eawag-PPS and hierarchical clustering to generate four
249 dendrograms that cluster micropollutants together based on different assumptions about their
250 biotransformations. For example, prediction method (2) yields a matrix of all possible predicted
251 aerobic biotransformations for each micropollutant (*i.e.*, no relative reasoning, neutral likelihood)
252 and, as a consequence, clusters micropollutants based on overall structural similarity. Conversely,
253 prediction method (3) yields a matrix of only likely aerobic biotransformations, and may exclude
254 some predictions based on the presence of another, particularly labile functional group; the
255 resulting clusters are expected to comprise groups of micropollutants that all contain a single,
256 particularly labile functional group. The dendrogram resulting from prediction method (1) is
257 provided in **Figure 3** and dendrograms resulting from the other prediction methods are provided
258 in **Figures S2-S4**. To evaluate the four prediction methods and to identify the dendrogram that
259 yields the most plausible micropollutant clusters, we examined each dendrogram in the context

of known micropollutant biotransformations that have been observed in full-scale WWTPs or in bioreactors seeded with wastewater microbial communities. We identified literature reported biotransformation pathways for 15 of the 84 micropollutants in our database; the reported biotransformation products and the corresponding btrules for each micropollutant are provided in **Table S3**. Prediction methods (3) and (4) both relied on the likely likelihood cutoff. This stringent criterion resulted in a number of micropollutants with no predicted biotransformations (see **Figure S3** and **S4**). Consequently, the dendrograms resulting from these prediction methods are dominated by large clusters of micropollutants with no predicted biotransformations and do not provide meaningful micropollutant clusters. Prediction methods (1) and (2) rely on the neutral likelihood cutoff, and explore the effects of relative reasoning on micropollutant clusters. Although the dendrograms resulting from both prediction methods yield plausible clusters of micropollutants, **Table S3** confirms that prediction method (1) generates more micropollutant clusters that contain micropollutants that undergo the same known biotransformation during activated sludge treatment. For example, prediction method (1) generates a micropollutant cluster that contains SMX-acetate and acetaminophen, two micropollutants that undergo amide hydrolysis during activated sludge treatment.[53,54] Additionally, the micropollutants contained in clusters generated from prediction method (1) are generally more closely related than those resulting from prediction method (2); this is evidenced by the height of the clusters which is reflected in the y-axis values of the dendrogram. For these reasons, we selected prediction method (1) to support our downstream analyses.

3.3 Evaluating R% values among individual seed micropollutants

We previously developed and validated a multivariable framework to identify the drivers of aerobic biodegradation in soil systems for individual micropollutants.[38] This approach relied

283 on at least 14 independent observations of aerobic biodegradation of an individual micropollutant,
284 along with associated environmental parameters (*e.g.*, organic carbon content of soil). Therefore,
285 we examined our dataset to identify those individual micropollutants for which we had more than
286 14 independent observations of *R%* and defined these as "seed" micropollutants. The seed
287 micropollutants were ibuprofen (n=34), naproxen (n=30), sulfamethoxazole (n=26),
288 acetaminophen (n=25), mefenamic acid (n=20), diclofenac (n=19), trimethoprim (n=17). We
289 performed a GAM and stepwise regression on *R%* as a response variable with the six process
290 parameters as predictor variables for each of the ten imputed datasets for each of the seed
291 micropollutants. Robust models were generated for trimethoprim and mefenamic acid which
292 exhibited adjusted R^2 values of 0.49-0.51 and 0.89-0.99, respectively. The remaining seed
293 micropollutants exhibited wider ranges of adjusted R^2 values, including very low adjusted R^2
294 values. We attribute these inferior results to the rather narrow ranges of reported *R%* values for
295 those micropollutants or at least one relatively poor imputation. Nevertheless, the adjusted R^2
296 value for the best imputation among the seed micropollutants is relatively high (compared to the
297 full dataset provided in **Figure 2**), except for diclofenac (0.08). The poor performance of the
298 models for diclofenac could be the result of limited data or because the important factors driving
299 the *R%* are not included among our predictors.

300 *3.4 Evaluating R% values among micropollutant clusters*

301 To test the hypothesis that accurate and robust predictions of micropollutant removal
302 during wastewater treatment can be achieved for micropollutants that undergo the same types of
303 enzyme-catalyzed biotransformations, we located each of the seed micropollutants in the
304 dendrogram provided in **Figure 3** and built micropollutant clusters based on each localized "leaf"
305 in the dendrogram. We defined a leaf as a cluster of micropollutants that contains at least one

306 seed micropollutant and within which all micropollutants are connected at a height of less than 5
307 (arbitrary units) in the dendrogram. We identified six micropollutant clusters in this way. The six
308 micropollutant clusters contain 15 micropollutants and 207 independent observations, which
309 covers 39% of our complete dataset.

310 The acetaminophen (ACE) cluster includes acetaminophen and SMX-acetate which both
311 reportedly undergo amide hydrolysis (btrule 67) during activated sludge treatment.[30] Although
312 bezafibrate clustered outside of the ACE leaf, we evaluated the ACE cluster with and without
313 bezafibrate because it is also known to undergo amide hydrolysis during activated sludge
314 treatment.[54] The sulfamethoxazole (SMX) cluster includes sulfamethazine, sulfadiazine,
315 sulfamethoxazole, and sulfamerazine, all of which contain a sulfonamide functional group and
316 all were reported to undergo sulfonamide hydrolysis (btrule 144) in activated sludge, which is
317 consistent with observations in experiments with bacteria isolated from activated sludge.[55–57]
318 Diclofenac (DCL) and mefenamic acid form their own cluster; these two micropollutants are the
319 only micropollutants in the database that contain a 1-amino-2-unsubstituted aromatic which is
320 what drives the clustering. The Eawag-PPS predicts dihydroxylation of the unsubstituted
321 aromatic and cleavage of the amino group (btrule 65). This biotransformation was reported for
322 diclofenac in activated sludge,[58] though, to the best of our knowledge, no biotransformation
323 pathway has been reported for mefenamic acid. Nevertheless, the similar structures of DCL and
324 mefenamic acid suggest similar biotransformation pathways. The naproxen (NPX) cluster
325 includes sulfamonomethoxine, and the clustering is driven by the shared presence of a methyl-
326 substituted ether group. Ether demethylation (btrule 23) is a known biotransformation pathway
327 for naproxen during wastewater treatment.[59,60] No known biotransformation of
328 sulfamonomethoxine has been reported in activated sludge, though one could anticipate

sulfonamide hydrolysis (btrule 144) or ether demethylation (btrule 23) as likely biotransformations. The trimethoprim (TRI) cluster contains tris(2-chloroethyl)phosphate and 4-isobutylacetophenone; all three micropollutants exclusively contain secondary aliphatic carbons that can be oxidized to alcohols (btrule 241), which is the feature that drives their clustering. Hydroxylation products have been reported for trimethoprim and tris(2-chloroethyl)phosphate,[61,62] though no literature was identified on the biotransformation of 4-isobutylacetophenone. Finally, ibuprofen (IBP) was clustered exclusively with lilial. These micropollutants exclusively contain primary or secondary carbons adjacent to unsaturated carbons (btrule 333, btrule 242), and hydroxylation of those functional groups has been reported for ibuprofen.[59]

Based on this review of predicted and observed biotransformations among these 15 micropollutants (16 when bezafibrate is included, and 221 independent observations), it is reasonable to expect that *R%* within each of the six micropollutant clusters is determined by the same set of predictors. We performed a GAM and stepwise regression on *R%* as a response variable with six physicochemical properties and six process parameters as predictor variables for each of the ten imputed datasets for each of the micropollutant clusters. The range of adjusted R^2 values among the ten imputed datasets for each cluster is provided in **Table 1**. Prediction plots for the imputed dataset for each cluster that yielded the highest adjusted R^2 value are presented in **Figure S5**. The ranges of the adjusted R^2 values for the micropollutant clusters were generally tighter compared to the ranges for the seed micropollutants, which suggests more robust regression results from the imputations for micropollutant clusters than seed micropollutants. **Figure S5** also demonstrates that each cluster yielded a greatly improved performance of GAM and stepwise predictions compared to the full dataset (**Figure 2**), which

indicates that the variability in *R%* among the micropollutants can be better explained by a similar set of physicochemical properties and process parameters. The model developed for the ACE+BEZ cluster (adjusted R^2 range 0.59-0.75) exhibited improved performance compared to the ACE cluster (adjusted R^2 range 0.25-0.52). This suggests that micropollutants that are known to undergo the same initial biotransformation are best clustered together, and shows that our prediction-based approach to micropollutant clustering can lead to some inaccuracies when other functional groups are present in the micropollutant. Based on this improved performance, we only consider the ACE+BEZ in our downstream analyses. Finally, it is notable that the model developed for the SMX cluster exhibited the poorest performance. This could be explained by alternative pathways being expressed under different conditions; it has been reported that some sulfonamide antibiotics can undergo a series of hydrolysis, oxidation, and decarboxylation reactions as part of the pterin-conjugate pathway.[57]

3.5 Stability selection and identification of important physicochemical properties

We then implemented a stability selection workflow to identify the key physicochemical properties and process parameters that determine the *R%* in each of the six micropollutant clusters. The importance of each predictor for determining the *R%* of micropollutants and the averaged adjusted R^2 values resulting from stability selection are provided in **Table 1**. Importantly, we found that *S*, *charge*, and *HLC* were never selected into well-performing models, which suggests that these physicochemical properties are not important in differentiating the variable *R%* among micropollutants in each of the clusters; this is not surprising considering that the micropollutants contained in each of the clusters are structurally similar and have similar values for these properties. Nevertheless, K_d, *MW*, and K_{ow} were selected into well-performing models for at least one micropollutant cluster. To further evaluate the ways in which each

predictor contributes to the variability in *R%* values for each micropollutant cluster, we provide the GAM stepwise shape functions in **Figure 4**.

K_d values were selected as an important predictor in at least 40% of the well-performing models for all micropollutant clusters except the TRI cluster. This is a particularly interesting finding because we only included those micropollutants with K_d values less than 500 $L \cdot kg^{-1}$ in our analysis as a means to isolate biotransformation as the dominant removal mechanism.[35,36] Our results suggest that the role that K_d plays in determining *R%* among micropollutants in each cluster is not so clear. For example, micropollutants in the ACE+BEZ, DCL, and NAP clusters exhibit a positive association between K_d values and *R%*. This could be explained by increased adsorption to sludge flocs for micropollutants with higher K_d values, despite our selection criterion that excluded micropollutants with K_d values greater than 500 $L \cdot kg^{-1}$.[63–65] This interpretation would suggest that our selected K_d threshold was not conservative enough to fully isolate biotransformation as the sole removal process; indeed other studies have suggested a more appropriate threshold could be as low as 300 $L \cdot kg^{-1}$.[66] Conversely, the IBP and SMX clusters exhibit a negative association between K_d values and *R%*. Others have also reported a similarly counterintuitive relationship between K_{oc} values and *R%*.[9] It is not clear why a negative association would exist between K_d values and *R%*, though it must be noted that the shape function for K_d in the IBP cluster is driven by a single value, which may bias the result.

MW was selected as an important predictor in at least 50% of the well-performing models for the ACE+BEZ and TRI clusters. The shape functions in **Figure 4** show a negative association between *MW* and *R%* which agrees with our expectation that larger molecules are more difficult to biotransform due to uptake limitations.[3] The micropollutants in the ACE+BEZ and TRI

clusters exhibit the largest differences in *MW* among the clusters, which could explain why *MW* was only selected as an important predictor for those clusters.

Finally, K_{ow} is selected as an important physicochemical property in differentiating *R%* in 18% of the well-performing models for the SMX cluster. However, the association between K_{ow} and *R%* in the shape function is not strong, so K_{ow} is not generally considered important in determining the *R%* of micropollutants included in this study.

3.6 Stability selection and identification of important process parameters

With respect to process parameters, we again see that most (if not all) are important in explaining the variability in *R%* among the micropollutants in each cluster. The size of the population that the plant serves (*P*) and the flow rate (*Q*) reflect information about the size of the WWTPs. Generally, increasing flow rate increases substrate loading and may increase biodiversity of the active biomass, resulting in a broader range of metabolic activity.[30] *P* and/or *Q* were selected as important predictors for all micropollutant clusters. The shape functions for *P* and *Q* provided in **Figure 4** generally exhibit a weak association with *R%*. The major exceptions are a positive association with *P* and *Q* for the IBP cluster and a negative association with *P* for the ACE+BEZ cluster. These findings suggest that substrate loading can have variable effects on the types of biotransformations performed during activated sludge treatment.

HRT measures the time available for biodegradation to occur in the reactor; an increased *HRT* is generally expected to increase the *R%* of micropollutants that are moderately biodegradable and have low sorption potential.[67] However, literature also reported that the relationship between *HRT* and *R%* is not always clear and negative associations between *HRT* and *R%* have been reported.[68,69] HRT was selected as an important predictor for all of the micropollutant clusters, and the shape functions provided in **Figure 4** show increasing HRT has

420 a positive influence on SMX, DCL, TRI, and IBP clusters, but negative influence on ACE+BEZ
421 and NPX clusters. This finding agrees with previous observations and suggests that the ways in
422 which HRT influences *R%* is dependent on the biotransformation.

423 *SRT* is a measure of the mean residence time of microorganisms in the reactor, and
424 corresponds to the growth rate of microorganisms. High *SRT* allows for enrichment of slow-
425 growing microorganisms that can increase the functional diversity of the microbial
426 community.[70,71] Higher SRT generally enhances biodegradation of typically moderately
427 biodegradable micropollutants.[3,67] The strong positive associations between *SRT* and *R%* for the
428 ACE+BEZ, TRI, and IBP clusters presented in **Figure 4** are consistent with this expectation and
429 with literature. For example, a higher SRT allows for the growth of nitrifying bacteria which was
430 reported to increase trimethoprim removal during activated sludge treatment.[72] However, it has
431 also been reported that the effect of an increase in *SRT* is not always clear and may vary
432 significantly depending on the tested compounds.[21,66,68] We observed a weak association between
433 *SRT* and *R%* for micropollutants in the SMX and DCL clusters, and even a strong negative
434 association between *SRT* and *R%* for micropollutants in the NAP cluster. Previous studies have
435 also reported weak or negative associations between SRT and removal of sulfamethoxazole,
436 diclofenac, and naproxen[10,21,58,73] which supports our findings.

437 C_{inf} is always selected as an important predictor as initial concentrations of certain
438 micropollutants could affect biodegradation rates and initial transformation pathways. The shape
439 functions provided in **Figure 4** generally show a positive or weak association between C_{inf} and
440 *R%*. This finding agrees with literature that shows that higher concentrations of micropollutants
441 can result in greater removal.[74] The major exception is the TRI cluster, which exhibits a strong
442 negative association between C_{inf} and *R%*. Higher trimethoprim concentrations have been shown

443 to decrease trimethoprim biodegradation and lead to hydroxylation as the dominant initial
444 biotransformation pathway, while lower trimethoprim initial concentrations resulted in
445 alternative pathways being expressed.[61] The negative association exhibited among
446 micropollutants in the TRI cluster could be explained by a change in initial biotransformation at
447 varying concentrations.

448 *NDN* is a measure of the dissolved oxygen levels in the reactor. Dissolved oxygen levels
449 impact the physiology of the microorganisms involved in micropollutant biotransformation and
450 can play a pivotal role in controlling the biodegradation rate of certain micropollutants.[75] Aerobic
451 biotransformation could be limited by oxygen availability, thereby reducing micropollutant
452 biodegradation rates.[75] *NDN* was found to be an important predictor for the SMX, TRI, and IBP
453 clusters. The shape functions in **Figure 4** show that oxygen levels have a positive association
454 with *R%* for the SMX and TRI clusters, as suggested by the literature. However, a negative
455 association between oxygen levels and *R%* was noted for the IBP cluster. This finding suggests
456 that both aerobic and anoxic metabolisms may be important for micropollutants in the IBP
457 cluster which allows for other possible biodegradation pathways to occur simultaneously.[75]

458 **4. CONCLUSIONS**

459 This study aimed to identify the drivers of micropollutant removal in WWTPs. We
460 hypothesized that accurate and robust predictions of micropollutant removal during wastewater
461 treatment can be achieved only by isolating specific removal mechanisms and exploring them
462 independently for groups of micropollutants that are likely to undergo the same types of enzyme-
463 catalyzed biotransformations. Indeed, our results suggest that clustering micropollutants based on
464 their initial biotransformations can lead to improved prediction of micropollutant removal during
465 activated sludge treatment. We believe that the same approach of clustering micropollutants

466 based on their predicted initial biotransformations should also have broad utility in exploring the
467 drivers of micropollutant biotransformations in other environmental systems.

468 Further, the results of our stability selection approach provide novel insights on the types
469 of biotransformations performed by wastewater microbial communities and the factors that
470 influence those biotransformations. We found that K_d and *MW* were the most important
471 physicochemical properties and *SRT*, C_{inf}, and *NDN* were the most significant process parameters
472 that determine the variability in *R%* values among micropollutants that undergo the same types
473 of biotransformations. Importantly, the ways in which each of these predictors contributes to the
474 observed variability of *R%* is different among the micropollutants clusters. This suggests that
475 process modifications that target the enhanced removal of certain micropollutants might further
476 inhibit the removal of other micropollutants. Future studies are needed that couple process-level
477 observations to changes in microbial community structure to further clarify the mechanisms that
478 support these observations.

479 **ACKNOWLEDGEMENTS**

480 This study was funded by the CEFIC Long-range initiative under project identifier LRI-ECO31.
481 We appreciate the statistical advice from Dr. Deniz Akdemir from the Cornell University
482 Statistical Consulting Unit.

483

LIST OF FIGURES

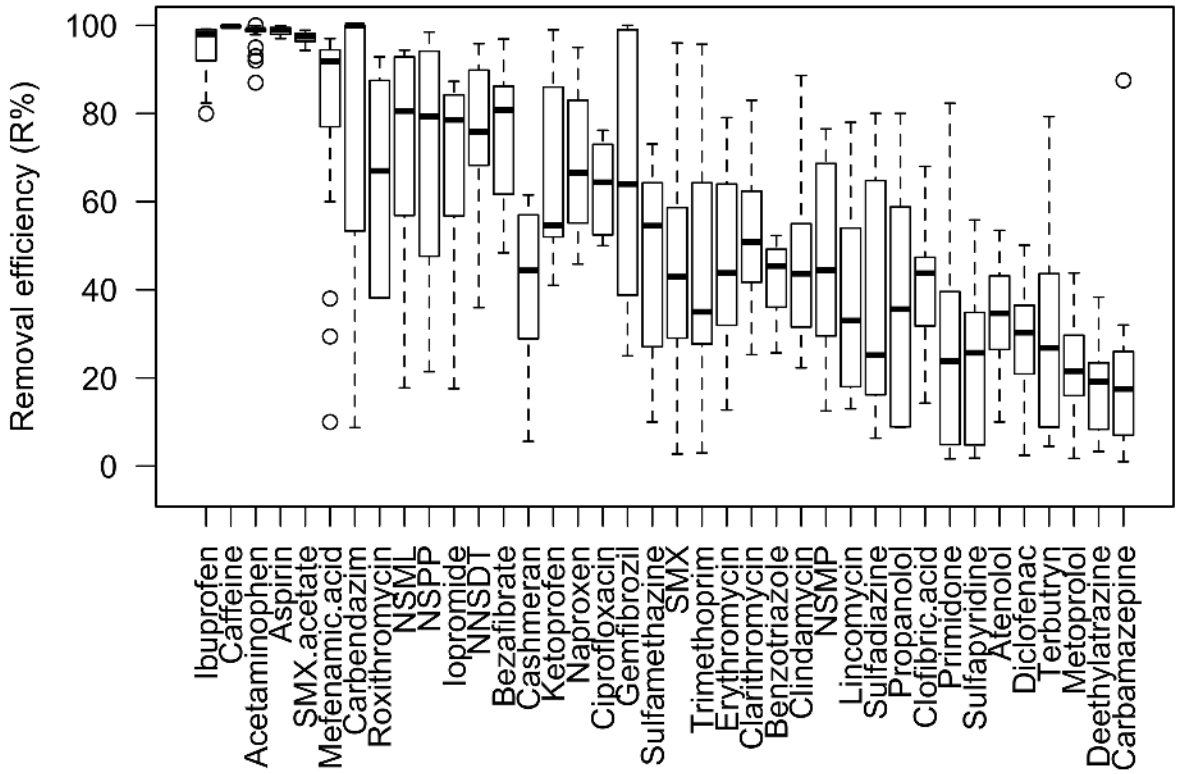
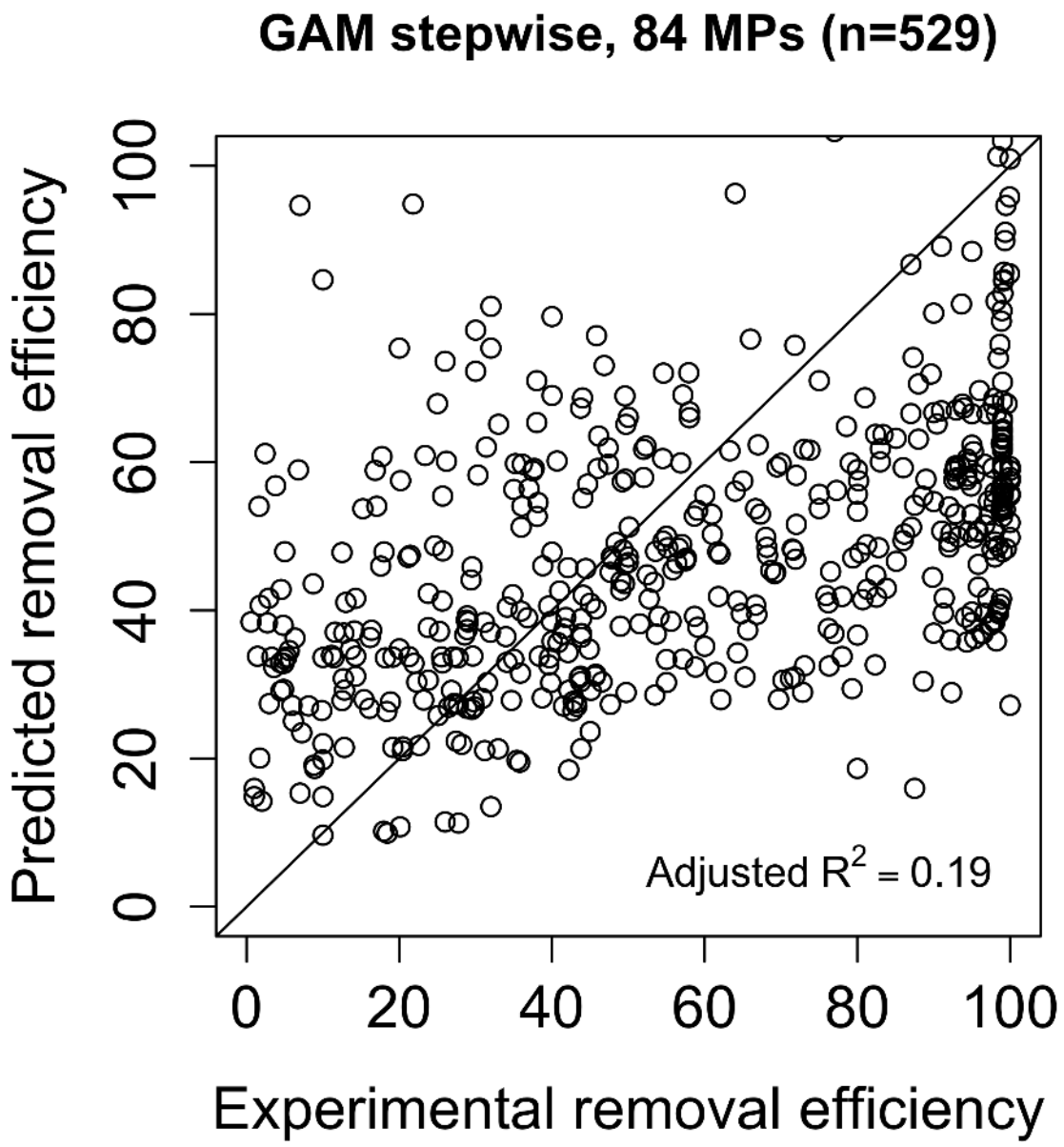


Figure 1. Boxplots of removal efficiency (*R%*) for selected micropollutants with more than four independent observations. SMX is sulfamethoxazole; SMX-acetate is N4-Acetylsulfamethoxazole; NSML is N-Nitrosodimethylamine; NSPP is N-Nitrosopiperidine; NSMP is N-Nitrosomorpholine; NNSDT is Nitrosodiethylamine.



494
495
496 **Figure 2**. Comparison of experimental removal efficiency and predicted removal efficiency of
497 the complete dataset with 84 micropollutants (n=529) using GAM and stepwise modeling.

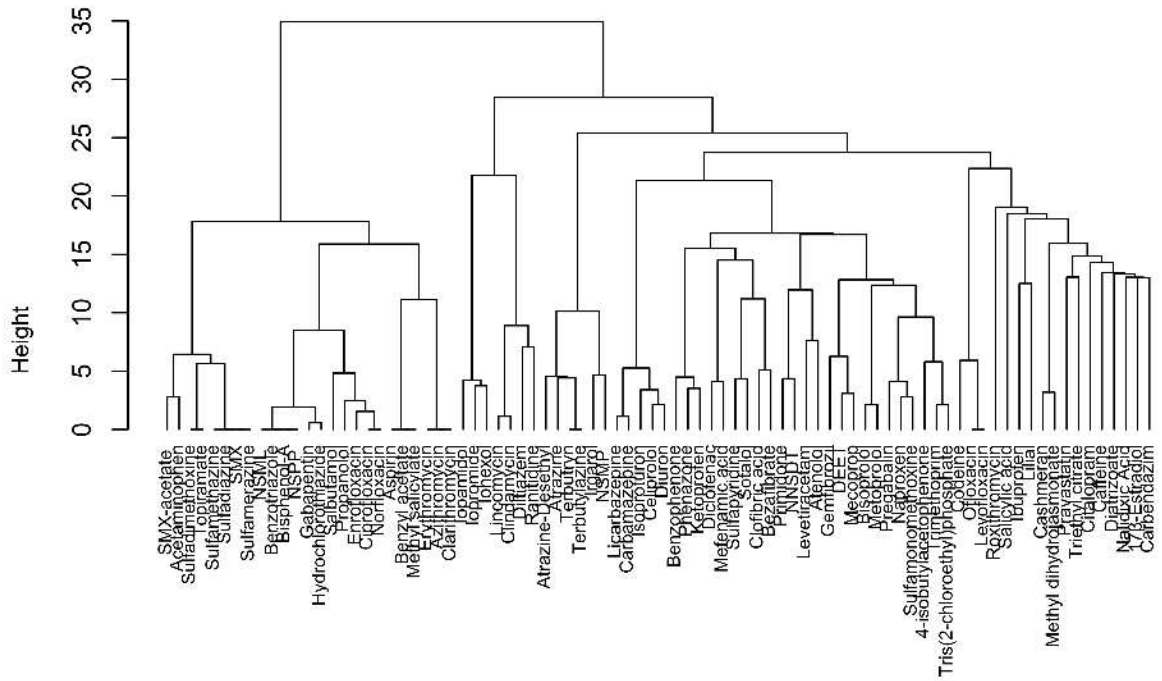


Figure 3. Hierarchical clustering of micropollutants based on their predicted initial biotransformation reactions using prediction method (1) relative reasoning, neutral likelihood cutoff in Eawag-PPS. The dendrogram was cut at height equaled 20 and yielded 12 major clusters. SMX is sulfamethoxazole; SMX-acetate is N4-Acetylsulfamethoxazole; NSML is N-Nitrosodimethylamine; NSPP is N-Nitrosopiperidine; NSMP is N-Nitrosomorpholine; NNSDT is Nitrosodiethylamine.

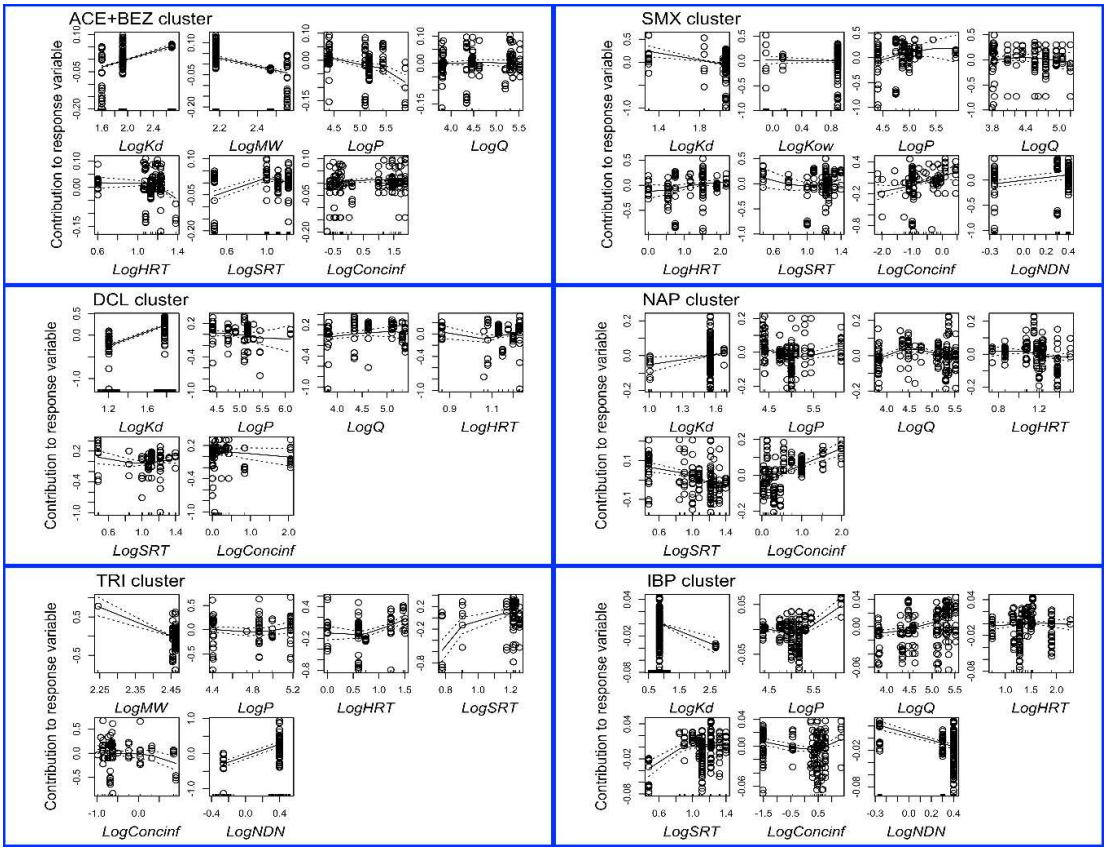


Figure 4. GAM stepwise shape functions describing the relationships between micropollutants removal efficiencies and each of the predictor variables in the final models for each cluster. The vertical axes indicate the contribution of each variable to the value of *R%*. Points represent collected data and partial residuals. Solid line represents the model fit, and dotted line represent confidence intervals of ±1 standard error.

TABLES

Table 1. Stability selection for the micropollutant clusters.

	ACE + BEZ	SMX	DCL	NAP	TRI	IBP
n obs	50	38	39	31	21	36
[a]range adj R^2	0.59-0.75	0.09-0.15	0.39-0.47	0.32-0.69	0.33-0.54	0.25-0.52
[b]avg adj R^2	0.71	0.35	0.58	0.65	0.73	0.73
K_d	98[c]	46	98	45	NA	50
MW	58	NA	NA	NA	52	NA
K_{ow}	NA	18	NA	NA	NA	NA
P	56	57	76	83	91	83
Q	50	41	28	75	NA	94
HRT	73	44	66	81	83	88
SRT	69	43	78	77	78	82
C_{inf}	90	49	69	74	84	69
NDN	NA	57	NA	NA	68	93

[a]Range of adjusted R^2 from GAM stepwise for the ten imputed datasets; [b]Average adjusted R^2 among all 10,000 models resulting from stability selection and VIF analysis; [c]Frequency (%) at which each physicochemical property and process parameter were selected into all 10,000 models at a significance level of <0.01; NA indicates the parameter was not included in final model resulting from stability selection and VIF analysis.

REFERENCES:

1 R. Oulton, T. Kohn and D. Cwiertny, Pharmaceuticals and personal care products in effluent matrices: A survey of transformation and removal during wastewater treatment and implications for wastewater management, *J. Environ. Monit.*, 2010, **12**, 1956–1978.

2 R. Eggen, J. Hollender, A. Joss, M. Schärer and C. Stamm, Reducing the Discharge of Micropollutants in the Aquatic Environment: The Bene fi ts of Upgrading Wastewater Treatment Plants, *Environ. Sci. Technol.*, 2014, **48**, 7683–7689.

3 Y. Luo, W. Guo, H. Ngo, H. H. Ngo, L. D. Nghiem and I. Hai, Faisal, A review on the occurrence of micropollutants in the aquatic environment and their fate and removal during wastewater treatment, *Sci. Total Environ.*, 2014, **473**–**474**, 619–641.

4 A. Joss, S. Zabczynski, A. Göbel, B. Hoffmann, D. Löffler, C. McArdell, T. Ternes, A. Thomsen and H. Siegrist, Biological degradation of pharmaceuticals in municipal wastewater treatment: Proposing a classification scheme, *Water Res.*, 2006, **40**, 1686–1696.

5 D. E. Helbling, D. R. Johnson, M. Honti and K. Fenner, Micropollutant biotransformation kinetics associate with WWTP process parameters and microbial community characteristics, *Environ. Sci. Technol.*, 2012, **46**, 10579–10588.

6 R. Gulde, D. E. Helbling, A. Scheidegger and K. Fenner, pH-dependent biotransformation of ionizable organic micropollutants in activated sludge., *Environ. Sci. Technol.*, 2014, **48**, 13760–13768.

7 B. Blair, A. Nikolaus, C. Hedman, R. Klaper and T. Grundl, Evaluating the degradation , sorption , and negative mass balances of pharmaceuticals and personal care products during wastewater treatment, *Chemosphere*, 2015, **134**, 395–401.

8 L. Stadler, L. Su, C. Moline, A. Ernstoff, D. Aga and N. Love, Effect of redox conditions on pharmaceutical loss during biological wastewater treatment using sequencing batch reactors, *J. Hazard. Mater.*, 2015, **282**, 106–115.

9 M. Douziech, I. R. Conesa, A. Benîtez-López, A. Franco, M. Huijbregts and R. van Zelm, Quantifying variability in removal efficiencies of chemicals in activated sludge wastewater treatment plants – a meta-analytical approach, *Environ. Sci. Process Impacts*, 2017, **20**, 171–182.

10 P. Falås, A. Wick, S. Castronovo, J. Habermacher, T. A. Ternes and A. Joss, Tracing the limits of organic micropollutant removal in biological wastewater treatment, *Water Res.*, 2016, **95**, 240–249.

11 K. M. Onesios, J. T. Yu and E. J. Bouwer, Biodegradation and removal of pharmaceuticals and personal care products in treatment systems: a review, *Biodegradation*, 2009, **20**, 441–466.

12 A. Ebele, M. A. Abdallah and S. Harrad, Pharmaceuticals and personal care products (PPCPs) in the freshwater aquatic environment, *Emerg. Contam.*, 2017, **3**, 1–16.

13 C. Miege, J. M. Choubert, L. Ribeiro, M. Eusebe, M. Coquery, C. Miege, J. M. Choubert, L. Ribeiro, M. Eusebe and M. C. Fate, Fate of pharmaceuticals and personal care products in wastewater treatment plants. Conception of a database and first results, *Environ. Pollut.*, 2009, **157**, 1721–1726.

14 P. Verlicchi, M. Aukidy and E. Zambello, Occurrence of pharmaceutical compounds in urban wastewater: Removal, mass load and environmental risk after a secondary treatment—A review, *Sci. Total Environ.*, 2012, **429**, 123–155.

15 M. Rojas, C. Leung, F. Bonk, Y. Zhu, L. Edwards, R. Arnold, A. Eduardo Sáez and G.

Klecka, Assessment of the Effectiveness of Secondary Wastewater Treatment Technologies to Remove Trace Chemicals of Emerging Concern, *Crit. Rev. Environ. Sci. Technol.*, 2013, **43**, 1281–1314.

16 C. Grandclément, I. Seyssiecq, A. Piram, C. Grandcl, P. Wong-wah-chung, G. Vanot, N. Tiliacos, N. Roche and P. Doumenq, From the conventional biological wastewater treatment to hybrid processes , the evaluation of organic micropollutant removal : A review, *Water Res.*, 2017, **111**, 297–317.

17 M. Rashed, *Chapter 7, Adsorption Technique for the Removal of Organic Pollutants from Water and Wastewater, Organic Pollutants - Monitoring, Risk and Treatment*, 2013.

18 S. Melvin and F. Leusch, Removal of trace organic contaminants from domestic wastewater: A meta-analysis comparison of sewage treatment technologies, *Environ. Int.*, 2016, **92**–**93**, 183–188.

19 D. E. Helbling, J. Hollender, H.-P. Kohler and K. Fenner, Structure-Based Interpretation of Biotransformation Pathways of Amide-Containing Compounds in Sludge-Seeded Bioreactors, *Environ. Sci. Technol.*, 2010, **44**, 6628–6635.

20 R. Gulde, U. Meier, E. L. Schymanski, H. E. Kohler, D. E. Helbling, S. Derrer, D. Rentsch and K. Fenner, Systematic Exploration of Biotransformation Reactions of Amine-Containing Micropollutants in Activated Sludge, *Environ. Sci. Technol.*, 2016, **50**, 2908–2920.

21 S. Achermann, P. Falås, A. Joss, C. Mansfeldt, Y. Men, B. Vogler and K. Fenner, Trends in Micropollutant Biotransformation along a Solids Retention Time Gradient, *Environ. Sci. Technol.*, 2018, **52**, 11601–11611.

22 T. Gallé, C. Koehler, M. Plattes, D. Pittois, M. Bayerle, R. Carafa, A. Christen and J.

Hansen, Large-scale determination of micropollutant elimination from municipal wastewater by passive sampling gives new insights in governing parameters and degradation patterns, *Water Res.*, 2019, **160**, 380–393.

23 M. Majewsky, T. Gallé, M. Bayerle, R. Goel, K. Fischer and P. A. Vanrolleghem, Xenobiotic removal efficiencies in wastewater treatment plants: Residence time distributions as a guiding principle for sampling strategies, *Water Res.*, 2011, **45**, 6152–6162.

24 M. Majewsky, J. Farlin, M. Bayerle and T. Gallé, A case-study on the accuracy of mass balances for xenobiotics in full-scale wastewater treatment plants, *Environ. Sci. Process. Impacts*, 2013, **15**, 730–738.

25 M. M. Benjamin and D. F. Lawler, *Water Quality Engineering: Physical / Chemical Treatment Processes*, Wiley, 1st edn., 2013.

26 J. Hollender, S. G. Zimmermann, S. Koepke, M. Krauss, C. S. McArdell, C. Ort, H. Singer, U. von Gunten and H. Siegrist, Elimination of Organic Micropollutants in a Municipal Wastewater Treatment Plant Upgraded with a Full-Scale Post-Ozonation Followed by Sand Filtration, *Environ. Sci. Technol.*, 2009, **43**, 7862–7869.

27 ChemAxon, MarvinSketch (17.2.20), calculation module developed by ChemAxon, *http://www.chemaxon.com*.

28 J. Struijs, *Application of SimpleTreat 4.0 in European substance regulations*, Bilthoven, The Netherlands, 2014.

29 T. Ternes, M. Bonerz, N. Herrmann, B. Teiser and H. Andersen, Irrigation of treated wastewater in Braunschweig, Germany: An option to remove pharmaceuticals and musk fragrances, *Chemosphere*, 2007, **66**, 894–904.

30 A. Göbel, C. S. Mcardell, A. Joss, H. Siegrist and W. Giger, Fate of sulfonamides , macrolides , and trimethoprim in different wastewater treatment technologies, *Sci. Total Environ.*, 2007, **372**, 361–371.

31 L. Kovalova, H. Siegrist, H. Singer, A. Wittmer and C. McArdell, Hospital Wastewater Treatment by Membrane Bioreactor: Performance and Efficiency for Organic Micropollutant Elimination, *Environ. Sci. Technol.*, 2012, **46**, 1536–1545.

32 A. Rodayan, M. Majewsky and V. Yargeau, Impact of approach used to determine removal levels of drugs of abuse during wastewater treatment, *Sci. Total Environ.*, 2014, **487**, 731–739.

33 R. Salgado, R. Marques, J. P. Noronha, G. Carvalho, A. Oehmen and M. A. Reis, Assessing the removal of pharmaceuticals and personal care products in a full-scale activated sludge plant, *Env. Sci Pollut Res Int*, 2012, **19**, 1818–1827.

34 S. Monteiro and A. Boxall, Occurrence and fate of human pharmaceuticals in the environment, *Rev Env. Contam Toxicol*, 2010, **202**, 53–154.

35 T. Ternes, A. Joss and H. Siegrist, Peer Reviewed: Scrutinizing Pharmaceuticals and Personal Care Products in Wastewater Treatment, *Environ. Sci. Technol.*, 2004, **38**, 392A--399A.

36 J. Radjenovic, M. Petrovic and D. Barceló, Fate and distribution of pharmaceuticals in wastewater and sewage sludge of the conventional activated sludge (CAS) and advanced membrane bioreactor (MBR) treatment, *Water Res.*, 2009, **43**, 831–841.

37 T. A. Ternes, N. Herrmann, M. Bonerz, T. Knacker, H. Siegrist and A. Joss, A rapid method to measure the solid-water distribution coefficient (Kd) for pharmaceuticals and musk fragrances in sewage sludge, *Water Res.*, 2004, **38**, 4075–4084.

636 38 Y. Wang, A. Lai, D. Latino, K. Fenner and D. E. Helbling, Evaluating the environmental
637 parameters that determine aerobic biodegradation half-lives of pesticides in soil with a
638 multivariable approach, *Chemosphere*, 2018, **209**, 430–438.
639 39 C. Carpenter, L. Y. Wong, D. Gutema and D. E. Helbling, Fall Creek Monitoring Station:
640 Using Environmental Covariates To Predict Micropollutant Dynamics and Peak Events in
641 Surface Water Systems, *Environ. Sci. Technol.*
642 40 P. D. Allison, Multiple Imputation for Missing Data: A Cautionary Tale, *Sociol. Methods*
643 *Res.*, 2000, **28**, 301–309.
644 41 P. T. Hippel, Regression with Mising Y's: An improved Strategy for Analysing Multiple
645 Imputed Data, *Sociol. Methodol.*, 2007, **37**, 265--291.
646 42 E. A. Stuart, M. Azur, C. Frangakis and P. Leaf, Practice of Epidemiology Multiple
647 Imputation With Large Data Sets: A Case Study of the Children' s Mental Health
648 Initiative, *Pract. Epidemiol.*, 2009, **169**, 1133–1139.
649 43 S. van Buuren and K. Oudshoorn, MICE: Multivariate imputation by chained equations in
650 R, *J. Stat. Softw.*, 2011, **45**, 1–67.
651 44 Y. Dong and C. J. Peng, Principled missing data methods for researchers, *Springerplus*,
652 2013, **2**, 1–17.
653 45 K. Fenner, J. Gao, S. Kramer, L. Ellis and L. Wackett, Data-driven extraction of relative
654 reasoning rules to limit combinatorial explosion in biodegradation pathway prediction,
655 *Bioinformatics*, 2008, **24**, 2079–2085.
656 46 L. B. M. Ellis, J. Gao, K. Fenner and L. P. Wackett, The University of Minnesota pathway
657 prediction system: predicting metabolic logic, *Nucleic Acids Res.*, 2008, **36**, W427–W432.
658 47 C. Carpenter and D. Helbling, Widespread Micropollutant Monitoring in the Hudson

River Estuary Reveals Spatiotemporal Micropollutant Clusters and Their Sources, *Environ. Sci. Technol.*, 2018, **52**, 6187–6196.

48 N. Meinshausen and P. Bühlmann, Stability Selection, *Stat. Methodol.*, 2010, **72**, 417–473.

49 A. Beinrucker, Ü. Dogan and G. Blanchard, Extensions of stability selection using subsamples of observations and covariates, *Stat. Comput.*, 2016, **26**, 1059–1077.

50 R. Shah and R. Samworth, Variable selection with error control: another look at stability selection, *J. R. Stat. Soc.*, 2013, **75**, 55–80.

51 Q. Long and B. Johnson, Variable selection in the presence of missing data: resampling and imputation, *Biostatistics*, 2015, **16**, 596–610.

52 R. O'brien, A Caution Regarding Rules of Thumb for Variance Inflation Factors, *Quality&Quantity*, 2007, **41**, 673–690.

53 A. Hart and D. Orr, The degradation of paracetamol (4-hydroxyacetanilide) and other substituted acetanilides by a Penicillium species, *Antonie Van Leeuwenhoek*, 1975, **41**, 239–247.

54 D. Krah, A. Ghattas, A. Wick, K. Bröder and T. Ternes, Micropollutant degradation via extracted native enzymes from activated sludge, *Water Res.*, 2016, **95**, 348–360.

55 A. Reis, M. Cvanvarova, Y. Liu, M. Lenz, T. Hettich, B. Kolvenback, P. Corvini and O. Nunes, Biodegradation of sulfamethoxazole by a bacterial consortium of Achromobacter denitrificans PR1 and Leucobacter sp. GP, *Environ. Biotechnol.*, 2018, **102**, 10299–10314.

56 N. Yang, J. Wan, S. Zhao and Y. Wang, Removal of concentrated sulfamethazine by acclimatized aerobic sludge and possible metabolic products, *PeerJ*, 2015, **3**, e1359.

57 S. Achermann, V. Bianco, C. Mansfeldt, B. Vogler, B. Kolvenbach, P. Corvini and K. Fenner, Biotransformation of Sulfonamide Antibiotics in Activated Sludge: The

Formation of Pterin-Conjugates Leads to Sustained Risk, *Environ. Sci. Technol.*, 2018, **52**, 6265–6274.

58 K. Jewell, P. Falås, A. Wick, A. Joss and T. Ternes, Transformation of diclofenac in hybrid biofilm–activated sludge processes, *Water Res.*, 2016, **15**, 559–567.

59 J. Quintana, S. Weiss and T. Reemtsma, Pathways and metabolites of microbial degradation of selected acidic pharmaceutical and their occurrence in municipal wastewater treated by a membrane bioreactor, *Water Res.*, 2005, **39**, 2654–2664.

60 S. Wolfson, A. Porter, J. Campbell and L. Young, Naproxen Is Transformed Via Acetogenesis and Syntrophic Acetate Oxidation by a Methanogenic Wastewater Consortium, *Environ. Microbiolgy*, 2018, **76**, 362–371.

61 K. Jewell, S. Castronovo, A. Wick, P. Falas, A. Joss and T. Ternes, New insights into the transformation of trimethoprim during biological wastewater treatment, *Water Res.*, 2016, **88**, 550–557.

62 USEPA, *TSCA Work Plan Chemical Problem Formulation and Initial Assessment, Chlorinated Phosphate Ester Cluster Flame Retardants, EPA Document# 740-R1-5001*, 2015.

63 A. Besha, A. Gebreyohannes, R. Tufa, D. Bekele, E. Curcio and L. Giorno, Removal of emerging micropollutants by activated sludge process and membrane bioreactors and the effects of micropollutants on membrane fouling: A review, *J. Environ. Chem. Eng.*, 2017, **5**, 2395–2414.

64 J. Virkutyte, R. Varma and V. Jegatheesan, *Treatment of Micropollutants in Water and Wastewater*, IWA publishing, London, 2010.

65 H. Rogers, Sources, behaviour and fate of organic contaminants during sewage treatment

and in sewage sludges, *Sci. Total Environ.*, 1996, **185**, 3–26.

66 A. Joss, E. Keller, A. C. Alder, A. Go, C. S. Mcardell, T. Ternes and H. Siegrist, Removal of pharmaceuticals and fragrances in biological wastewater treatment, *Water Res.*, 2005, **39**, 3139–3152.

67 P. Krzeminski, M. Tomei, P. Karaolia, A. Langenhoff, C. M. Almeida, E. Felis, F. Gritten, H. Andresen, T. Fernandes, C. Manaia, L. Rizzo and D. Fatta-Kassinos, Performance of secondary wastewater treatment methods for the removal of contaminants of emerging concern implicated in crop uptake and antibiotic resistance spread: A review, *Sci. Total Environ.*, 2019, **648**, 1052–1081.

68 N. Vieno, T. Tuhkanen and L. Kronberg, Elimination of pharmaceuticals in sewage treatment plants in Finland, *Water Res.*, 2007, **41**, 1001–1012.

69 M. Gros, M. Petrovic, A. Ginebreda and D. Barcelo, Removal of pharmaceuticals during wastewater treatment and environmental risk assessment using hazard indexes, *Environ. Int.*, 2010, **36**, 15–26.

70 C. Mansfeldt, S. Achermann, Y. Men, J. C. Walser, K. Villez, A. Joss, D. R. Johnson and K. Fenner, Microbial residence time is a controlling parameter of the taxonomic composition and functional profile of microbial communities, *ISME J.*, 2019, **13**, 1589–1601.

71 D. C. Vuono, J. Benecke, J. Henkel, W. C. Navidi, T. Y. Cath, J. Munakata-Marr, J. R. Spear and J. E. Drewes, Disturbance and temporal partitioning of the activated sludge metacommunity, *ISME J.*, 2015, **9**, 425–435.

72 A. Batt, S. Kim and D. Aga, Comparison of the occurrence of antibiotics in four full-scale wastewater treatment plants with varying designs and operations, *Chemosphere*, 2007, **68**,

428–435.

73 J. L. Santos, I. Aparicio, M. Callejon and E. Alonso, Occurrence of pharmaceutically active compounds during 1-year period in wastewaters from four wastewater treatment plants in Seville (Spain), *J. Hazard. Mater.*, 2009, **164**, 1509–1516.

74 K. Onesios-Barry, D. Berry, J. Proescher, I. K. Sivakumar and E. Bouwer, Removal of Pharmaceuticals and Personal Care Products during Water Recycling: Microbial Community Structure and Effects of Substrate Concentration, *Appl. Environ. Microbiol.*, 2014, **80**, 2440–2450.

75 L. Stadler and N. Love, Oxygen Half-Saturation Constants for Pharmaceuticals in Activated Sludge and Microbial Community Activity under Varied Oxygen Levels, *Environ. Sci. Technol.*, 2019, **53**, 1918–1927.